Paper

Impact of the duration of speech sequences on speech quality

Peter Počta and Martin Vaculík

Abstract—This paper describes simulations of speech sequences transmission for intrusive measurement of voice transmission quality of service (VTQoS) in the environment of IP networks. The aim of the simulations was to investigate the impact of the different durations of speech sequences on speech quality from the jitter rate and packet loss point of view in IP networks. The ITU-T G.729 and ITU-T G.723.1 encoding schemes were used for the purpose of the simulations. The assessment of speech quality was realized by means of perceptual evaluation of speech quality (PESQ) algorithm. A comparison of the impact of different durations of speech sequences on speech quality and determination of the optimal duration of speech sequence for measurements of speech quality in telecommunication networks, is the aim of this paper.

Keywords— voice transmission quality of service, speech sequence, intrusive measurement, perceptual evaluation of speech quality, jitter rate, packet loss.

1. Introduction

Voice transmission quality of service (VTQoS) is one of the important parts of quality of service (QoS). It is very important for providers as well as for users. When communication networks incorporate more and more transmission technologies, an increase in complication and the complexity of networks is seen. Measurement of the voice transmission quality becomes only platform that is available for simultaneous comparison of different transmission technologies and that is the most relevant to the view of the users. Of course, it is possible to measure and evaluate the transmission parameters of the networks, but only the evaluation of end-to-end quality provides optimal results because of the complexity of network technologies. Thus, it is the evaluation by similar way as users do. Since voice service is the most wide-spread service, in which a user uses filter and predicative abilities of human brain, it is crucial to optimally evaluate a quality of such service.

Evaluation of a quality of the voice service may be performed using intrusive or non-intrusive methods, objectively or subjectively. Using non-intrusive method, we only monitor existing dialogue. The drawback of this method is that the evaluation algorithm cannot utilize an original sample of the primary signal. Thus, it is very difficult to detect some types of signal distortion that occure during transmission. In the intrusive methods, only a test voice sample is transmitted. These methods have been known since the beginning of the telecommunication technologies, when the special sequences of vowels (known as logathoms) were transmitted after the connection had been built-up. A receiver had to recognize these logathoms. This way of subjective evaluation has been used till nowadays (e.g., absolute category rating method).

Today's technical and software facilities provide an objectification of this measurement method by transmitting the voice sample defined beforehand, its receiving on the destination side, and a comparison of the transmission sample and the original sample using the suitable algorithm that imitates the way of perception and evaluation of the quality transmission opinion by an average listener. It is for example perceptual speech quality measurement (PSQM) algorithm defined in ITU-T Rec. P.861 [6] or perceptual evaluation of speech quality (PESQ) algorithm defined in ITU-T Rec. P.862 [7]. The PSQM algorithm is based on comparison of the power spectrum of the corresponding sections of the original and the received signals. The results of this algorithm more correlate with the results of listening tests, in comparison with E-model [1]. At present, this algorithm is no longer used because of a raw time alignment. Instead of it the PESQ algorithm is used. The PESQ algorithm facilitates with very fine time alignment and one single interruption are also taken into account in the calculation of mean opinion score (MOS). It is possible to use PESQ in mobile networks as well as in networks based on packet transmission. The disadvantages include impossibility to use it for codec with data rate lower than 4 kbit/s and higher calculation load what is caused by recursions in the algorithm.

The ITU-T Rec. P.862.3 [10] recommends to use the speech sequence in duration in the range from 8 s till 30 s for the purpose of speech quality measurement. Here we focus on the impact of different durations of speech sequences on speech quality from the point of view of jitter rate and packet loss in IP network. We want to compare two durations of speech sequences from the range that is recommend for purpose of speech quality measurements in the ITU-T Rec. P.862.3 [10]. The speech sequences in duration of 10 and 30 s are compared in this paper. The rest of paper is organized as follows: Section 2 describes simulation design. Section 3 presents the simulation results. Section 4 concludes the paper and suggests some future studies.

2. Simulation design

2.1. Experimental design

The transmission simulations were carried out on Gaoresearch's (freeware) online simulator [11]. The simulation model of transmission chain with codec ITU-T G.723.1 [2]

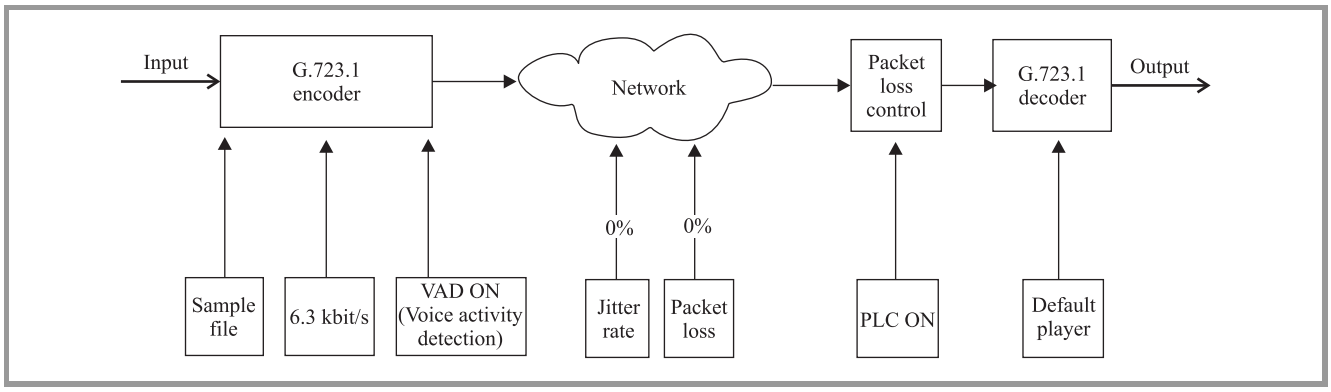

Fig. 1. The simulation model of transmission chain with codec ITU-T G.723.1.

is depicted in Fig. 1. The simulation model enables to change jitter rate and packet loss parameters in the range from 0% to 10%. The simulation model renders voice activity detection (VAD) and packet loss control (PLC) functions.

The speech sequences described in Subsection 2.2 were used for the simulations. The simulations were performed for different setting of packet loss and jitter rate parameters and by using these 2 encoding schemes:

- ITU-T G.723.1 (6.3 kbit/s) [2],
- ITU-T G.729 [3].

The simulations of jitter influence were done for the values of jitter rate in the range from 0% to 10%. Jitter rate is defined as percentual number of the packets, whose jitter value is above maximum tolerated jitter for given connection. Jitter is a measure of variation in latency over time. Jitter is caused by random variation of the momentary traffic load. This simulator tolerates the jitter below 90 ms. The packets delivered after this time are further not processed, they are also dropped out. The packets loss influence was investigated for the values of packet loss in the range from 0% to 10%. The packet loss parameter is defined as the percentual number of the packets that were lost during the transmission. Packets may lost, due to high bit error rate of the transmission channel and high traffic load. The VAD and PLC functions were activated for all the performed simulations. Finally, MOS was measured by PESQ algorithm.

2.2. Description of speech sequences

The speech sequences selection should follow the criteria given by ITU-T Rec. P.830 [5] and ITU-T Rec. P.800 [4]. The speech sequence should include bursts separated by silence periods, and are normally 1–3 s in duration. Also it should be active for 40–80% of the length. The speech sequences are composed from speech records. In our experiments, these speech records come from a Slovak database. In each set, two female and two male speech utterances were used. The speech sequences was stored in 16-bit, 8000 Hz linear PCM, and were 10 s in duration with 61.5% of active speech interval and 30 s in duration with 57% of active speech interval.

2.3. Assessment of speech quality

The MOS was measured by PESQ [7] metric, the most recent ITU-T standard for objective speech quality assessment. The PESQ combines merits of PAMS and PSQM99 (an updated version PSQM), and adds new methods for transfer function equalization and averaging distortions over time. It can be used in wider range of network conditions, and gives higher correlation with subjective tests and the other objective algorithms [7, 12]. In contrast to the conversational model; PESQ is a listening-only model, the degraded sample is time-aligned with the reference sample during pre-processing. The PESQMOS values do not reflect the effects of delay on speech quality.

3. Simulation results

The simulation was independently performed 3 times under the same testing conditions. The MOS results were averaged out and the standard deviation was kept within 0.0894 MOS for speech sequences in duration of 10 s and 0.0664 MOS for speech sequences in duration of 30 s. The values of standard deviation for both durations and types of speech sequences, and for both encoding schemes are summarized in Table 1.

Table 1
Standard deviation values

Encoding scheme	10 s		30 s	
	M	F	M	F
G.729 (jitter)	0.0792	0.0894	0.0553	0.0385
G.729 (packet loss)	0.0553	0.0701	0.0403	0.0448
G.723.1 (jitter)	0.0458	0.0727	0.0328	0.0664
G.723.1 (packet loss)	0.0339	0.0561	0.0288	0.0375

3.1. The simulation results of jitter influence on speech sequences transmission

The graphs (Figs. 2–5) represent the dependence of MOS values change on the percentual number of packets, whose

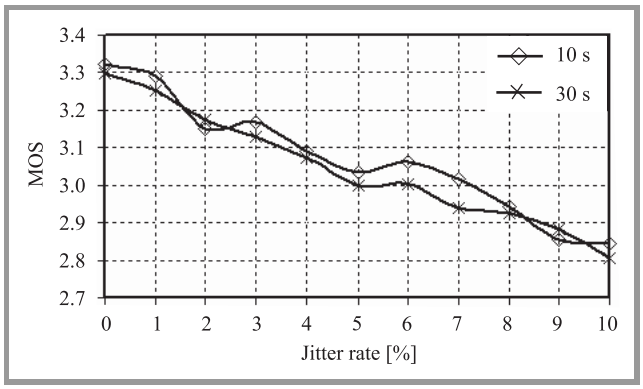

Fig. 2. Impact of jitter rate on speech quality of male speech sequences for ITU-T G.729.

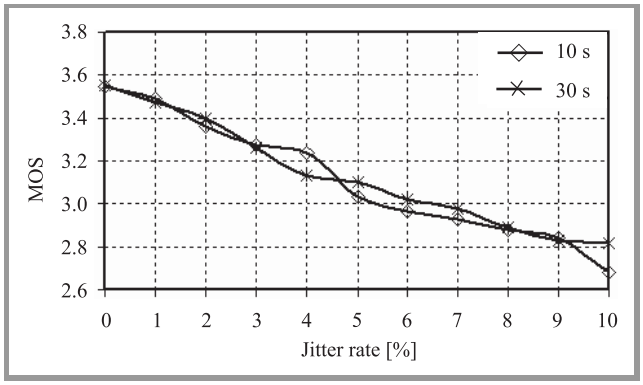

Fig. 3. Impact of jitter rate on speech quality of female speech sequences for ITU-T G.729.

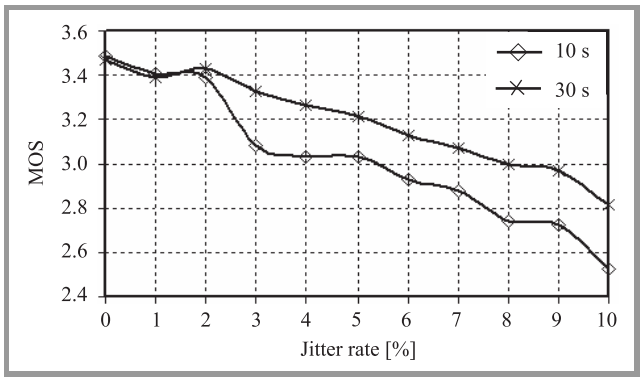

Fig. 4. Impact of jitter rate on speech quality of male speech sequences for ITU-T G.723.1 (6.3 kbit/s).

jitter overstepped the time of 90 ms. Every packet whose jitter overstepped the time of 90 ms is dropped. Non-uniform distribution of the dropped packets in both active speech and silence periods causes a smooth undulation

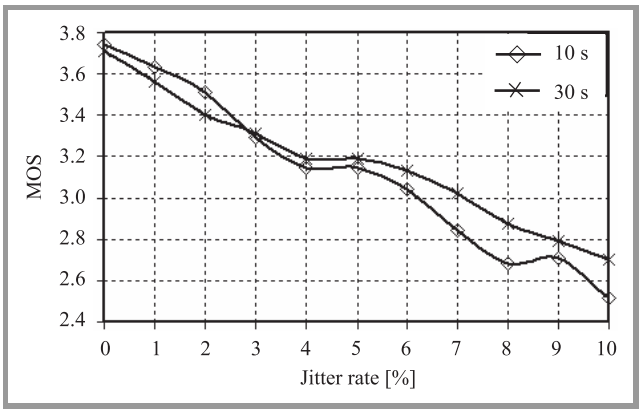

Fig. 5. Impact of jitter rate on speech quality of female speech sequences for ITU-T G.723.1 (6.3 kbit/s).

of the characteristics. In the case of zero jitter rate, MOS value change is cause only by encoding scheme. The graphs show only average values for female speech sequences and for male speech sequences, respectively.

3.2. The simulation results of packets loss influence on speech sequences transmission

The graphs (Figs. 6–9) represent the dependence of MOS values change on packets loss that means of percentual

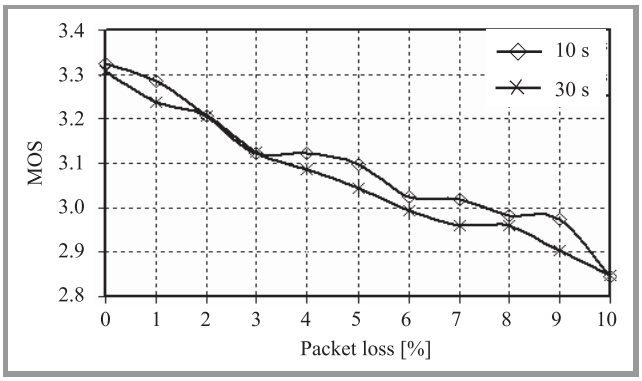

Fig. 6. Impact of packet loss on speech quality of male speech sequences for ITU-T G.729.

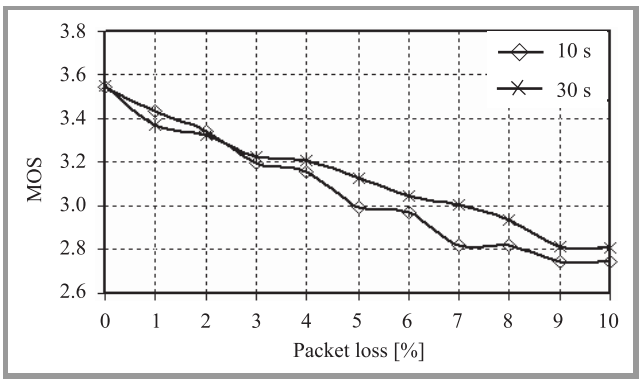

Fig. 7. Impact of packet loss on speech quality of female speech sequences for ITU-T G.729.

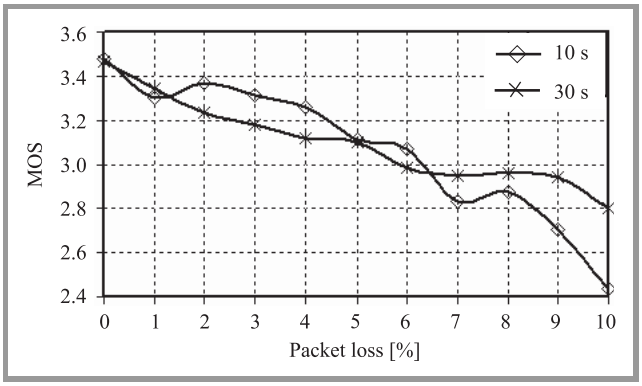

Fig. 8. Impact of packet loss on speech quality of male speech sequences for ITU-T G.723.1 (6.3 kbit/s).

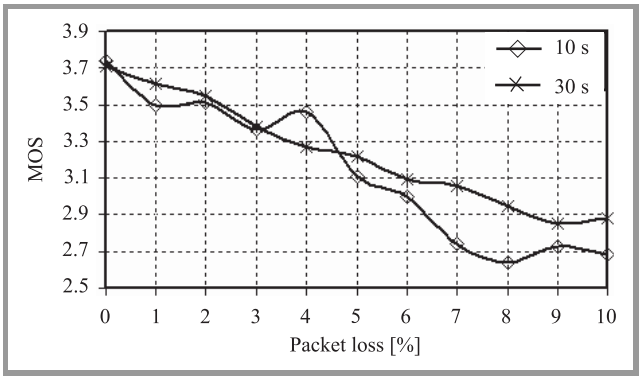

Fig. 9. Impact of packet loss on speech quality of female speech sequences for ITU-T G.723.1 (6.3 kbit/s).

number of the packets, which were not delivered. The smooth undulation of the characteristics is caused by non-uniform distribution of lost packets in both active speech and silence periods. In the case of zero packet loss, MOS value change is cause only by encoding scheme. The graphs show only average values for female speech sequences and for male speech sequences, respectively.

4. Conclusion

This paper has investigated the impact of the different durations of speech sequences for codecs ITU-T G.729 and ITU-T G.723.1 on speech quality from the jitter rate and packet loss point of view in IP networks. The results show that the difference in duration of speech sequences has the impact on speech quality. It is caused by more active speech and silence periods in sequences in duration of 30 s. A probability of impact of jitter rate and packet loss on speech quality is higher if more periods are available. It means that it is possible to capture more changes in speech quality in this case. The capture of more changes allows to realize the more precise measurements of speech quality in telecommunication networks. It was a first idea for the comparison of these two types of speech sequences from the duration point of view.

The speech sequences in duration of 30 s have the smoother characteristic than speech sequences in duration of 10 s. The values of the standard deviation (Table 1) are lesser for speech sequences in duration of 30 s than for speech sequences in 10 s duration. Therefore this type of speech sequences is better for speech quality measurements. The more precise results are obtained by means of longer speech sequences in 30 s duration.
In the future, this problem will be investigated in the converged network of the University of Žilina.

References

[1] "The E-model, a computational model for use in transmission planning", ITU-T Rec. G.107 (03/2005).

[2] "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s", ITU-T Rec. G.723.1 (05/2006).

[3] "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)", ITU-T Rec. G.729 (03/1996).

[4] "Methods for subjective determination of transmission quality", ITU-T Rec. P.800 (08/1996).

[5] "Subjective performance assessment of telephone-band and wide-band digital codecs", ITU-T Rec. P. 830 (02/1996).

[6] "Objective quality measurement of telephone-band (300–3400 Hz) speech codecs", ITU-T Rec. P.861 (02/1998).

[7] "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and peech codecs", ITU-T Rec. P.862 (02/2001).

[8] "Mapping function for transforming P.862 raw result scores to MOS-LQO", ITU-T Rec. P.862.1 (11/2003).

[9] "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs", ITU-T Rec. P.862.2 (11/2005).

[10] "Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2", ITU-T Rec. P.862.3 (11/2005).

[11] "Gaoresearch online simulator", http://www.gaoresearch.com/products/speechsoftware/speechsoftware.php

[12] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) a new method for speech quality assessment of telephone network and codecs", in *Proc. IEEE Int. Conf. Acoust., Speech Sig. Proces*., Salt Lake City, USA, 2001, pp. 749–752.

[13] L. Ding and R. A. Goubran, "Assessment of effects of packet loss on speech quality in VoIP", in *Proc. 2nd IEEE Int. Worksh. Hapt., Audio Visu. Env. Their Appl. 2003, HAVE 2003*, Ottawa, Canada, 2003, pp. 49–54.

[14] M. Varela, I. Marsh, and B. Gronvall, "A systematic study of PESQ's behavior", in *Proc. Conf. MESAQIN 2006*, Prague, Czech Republic, 2006.

[15] J. G. Beerends, E. Larsen, and N. Iyer, "Measurement of speech intelligibility based on the PESQ approach", in *Proc. Conf. MESAQIN 2004*, Prague, Czech Republic, 2004.

[16] J. Holub and A. Drozdová, "Proprietary low bit-rate radio-communication network – objective and subjective speech transmission quality assessment", in *Proc. Conf. MESAQIN 2006*, Prague, Czech Republic, 2006.

[17] W. A. Rix, "Comparison between subjective listening quality and P.862 PESQ score", in *Proc. Conf. MESAQIN 2003*, Prague, Czech Republic, 2003.

[18] J. Holub, R. Šmíd, and M. Bachtík, "Child listeners as the test subject – comparison with adults and P.862", in *Proc. Conf. MESAQIN 2003*, Prague, Czech Republic, 2003.

[19] S. Pennock, "Accuracy of the perceptual evaluation of speech quality (PESQ) algorithm", in *Proc. Conf. MESAQIN 2002*, Prague, Czech Republic, 2002.

Peter Počta was born in 1981, in Nové Zámky, Slovakia. He graduated from the University of Žilina, the Faculty of Electrical Engineering. He joined the Department of Telecommunications at the University of Žilina to obtain Ph.D. degree in telecommunications. His areas of interest include speech quality assessment, access networks, convergent networks, VoIP, VoWLAN, VoWiMAX.

e-mail: pocta@fel.uniza.sk
Faculty of Electrical Engineering
Department of Telecommunications
University of Žilina
Univerzitná st 1
010 26 Žilina, Slovakia

Martin Vaculík was born in 1951, in Žilina, Slovakia. He graduated from the University of Žilina as Dipl. Ing. in telecommunications in 1976. He received Ph.D. in 1987. The title of his thesis was "Some possibilities of dynamic routing control". He was with Siemens PSE Company, Optical Network Department in 2001–2002. Nowadays he works in the Department of Telecommunications at the University of Žilina as an Associate Professor. His areas of interest include switching, access networks and convergent networks. He is author of the book with the title "Access Networks" and co-author of 2 books: "ISDN" and "Broadband Networks".

e-mail: vaculik@fel.uniza.sk
Faculty of Electrical Engineering
Department of Telecommunications
University of Žilina
Univerzitná st 1
010 26 Žilina, Slovakia